

The Impact of Student Self-efficacy on Scientific Inquiry Skills: An Exploratory Investigation in *River City*, a Multi-user Virtual Environment

Diane Jass Ketelhut^{1,2}

This exploratory study investigated data-gathering behaviors exhibited by 100 seventh-grade students as they participated in a scientific inquiry-based curriculum project delivered by a multi-user virtual environment (MUVE). This research examined the relationship between students' self-efficacy on entry into the authentic scientific activity and the longitudinal data-gathering behaviors they employed while engaged in that process. Three waves of student behavior data were gathered from a server-side database that recorded all student activity in the MUVE; these data were analyzed using individual growth modeling. The study found that self-efficacy correlated with the number of data-gathering behaviors in which students initially engaged, with high self-efficacy students engaging in more data gathering than students with low self-efficacy. Also, the impact of student self-efficacy on rate of change in data gathering behavior differed by gender. However, by the end of their time in the MUVE, initial student self-efficacy no longer correlated with data gathering behaviors. In addition, students' level of self-efficacy did not affect how many different sources from which they chose to gather data. These results suggest that embedding science inquiry curricula in novel platforms like a MUVE might act as a catalyst for change in students' self-efficacy and learning processes.

KEY WORDS: gender; interactivity; longitudinal analysis; middle school; multi-user virtual environment; River City; scientific inquiry; self-efficacy.

INTRODUCTION

Many members of the business community, many policymakers, and much of the education community believe that science education is in crisis: Large numbers of students in our schools are not learning the science content or developing the appreciation of scientific inquiry needed to become the scientifically literate workers and citizens demanded by society (Business Higher Education Forum, 2005; Coble and Allen, 2005). Scholars dis-

agree about the reasons for this crisis, but most would concur that one cause is the lack of student engagement in science classes. According to the *Nation's Report Card* for 2005, only one-third of all students take the traditional 3 years of high school science (Chemistry, Biology and Physics), and nearly one-third only take a single year (Grigg *et al.*, 2006). Why is this?

While students' lack of engagement stems from multiple factors, this study focuses on one: students' sense of their own abilities to succeed in science. In his analysis of science education, Lemke (1990) concludes that, "Science is presented as being a *difficult* subject. When students fail to master it, they are encouraged to believe it is their own fault: they are just not smart enough to be scientists" (p. 138). In my

¹Temple University College of Education, Philadelphia, PA, USA

²To whom correspondence should be addressed;

E-mail: diane.jass.ketelhut@temple.edu

years of pre-college teaching, I found repeatedly that students who were otherwise confident and capable believed that they could not learn science, and so they chose not to engage in the activities and experimentation that are the heart of science. This perpetuated a vicious cycle: As the students lost confidence, they stopped trying to participate, which spawned further failure; as failures mounted, they lost even more confidence, and eventually opted out of learning science completely.

A recent longitudinal study by Tai *et al.*, reported in *Science* (2006) indicates one outcome of this cycle. These researchers discovered a strong relationship between eighth grade students who expected to have a career in science and those who eventually graduated with an undergraduate degree in science. Therefore, unless we can convince students early on in their schooling that science is achievable for them, we run the risk of perpetuating the idea that science is only for a very few elite students.

How do we break this pattern of failure? Some believe that reversing students' erosion of confidence, their loss of *self-efficacy*, is the place to start (Jinks and Morgan, 1996). Self-efficacy is the belief that one can succeed in performing particular behaviors; this has been shown to be more strongly related to academic outcomes than many other individual characteristics like student gender, student self-concept, or the perceived usefulness of the knowledge later in the student's life (Pajares and Miller, 1994). Stemming the loss of self-efficacy in science as students progress through school is a crucial first step in improving student educational outcomes in science. To stem this tide of decreasing confidence in science, we must better understand how differences in self-efficacy lead students to participate differentially in learning practices, which then ultimately result in divergent learning outcomes.

Therefore, I conducted an exploratory study to examine the relationship between students' self-efficacy on entry into authentic scientific activity and the growth of scientific inquiry behaviors they employed while engaged in that process over time. I did this using an innovative science curriculum delivered through a computer-based learning environment that records each student's conversations, movements, and activities while they are behaving as a practicing scientist in a "virtual world" called *River City*. Students were free to choose where in the virtual environment to explore, with whom to converse, what artifacts to examine, which data collection tools to

use, and what guidance to seek. In observing the students' scientific inquiry behaviors in this virtual world, I focused on the relationship between students' self-efficacy and how much scientific evidence they chose to gather, as well as the diversity of the types of scientific evidence they chose to utilize over time.

BACKGROUND AND CONTEXT

Scientific Inquiry and its Impact on Student Learning

For the last two decades, scientific inquiry has been a major curriculum standard in most policy doctrines (American Association for the Advancement of Science, 1990, 1993; National Research Council, 1996). The National Science Education Standards define scientific inquiry as

(Scientific) inquiry is a multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating the results. (National Research Council, 1996, p. 23).

This definition highlights the importance of making observations, formulating hypotheses, gathering and analyzing data, and forming conclusions from that data. A question underlying this work is whether participating in scientific inquiry improves student learning?

Several empirical studies have compared students studying in classrooms that promote scientific inquiry, as defined above, to those in other kinds of science classrooms. For example, Mason found no differences in achievement in college chemistry between students who took an inquiry-based chemistry course in high school and those who took a traditionally taught chemistry course (reported in Leonard *et al.*, 2001). However, Leonard *et al.* (2001) themselves found that students participating in a yearlong scientific inquiry-based biology course posted higher gains in biology concepts, and in the understanding of scientific processes. Furthermore, Alberts (2000) discovered that participating in scientific inquiry appears to improve retention of student learning. Before relating the above description and factors of scientific inquiry to the curricular context used in this study, I also overview the other core focus of this study: self-efficacy.

What is Known About Self-efficacy in Science and its Impact on Behavior

In seminal work, Bandura (1977) defined “self-efficacy” as the *belief* that one can *successfully perform* certain behaviors, such as graphing data. As such, self-efficacy is a belief in one’s abilities to accomplish a task, not a measure of those abilities. Pajares (1995, 2000) further argued that self-efficacy affects behavior by regulating an individual’s choices, the extent of his or her expended effort, and his or her emotional responses. In the classroom, students with higher self-efficacy are more likely to:

- Persevere in difficult situations (Lent *et al.*, 1984; Pajares, 2000);
- See complexity as a challenge (Pajares, 2000);
- Be engaged (Pajares, 2000);
- See failure as indication that more effort is needed (Bandura, 1986; Collins, 1984; Pajares, 2000);
- Choose specific strategies to enhance learning (Zimmerman and Bandura, 1994);
- Attribute success to ability (Pajares, 1995);

Students with lower self-efficacy are less likely to do the above, and more likely to:

- Equate failure to bad luck and poor ability (Pajares, 1995, 2000);
- Presume that a problem is more complex than it is (Pajares, 2000).

Not only does self-efficacy mediate behavior, but it also affects outcomes. Students with higher self-efficacy in a particular subject perform better and are more likely to be interested in a career in that field (Lopez and Lent, 1992; Pajares, 1997).

While I have described these relationships quite generally, there is much research suggesting that self-efficacy is context-dependent (Smith and Fouad, 1999), and scholars debate whether global measurements of self-efficacy are as strongly predictive of specific outcomes as their context-specific alternatives (Bandura, 1986; Bong, 1996; Pajares, 1996; Smith and Fouad, 1999). To measure self-efficacy in the learning of science in this study, I used a context-specific instrument to measure students’ self-efficacy for scientific inquiry, a measure that is specifically appropriate to the scientific tasks and problems posed to students in the *River City* project. In previous research (Ketelhut, 2004), I designed, piloted, and refined this measure, a subscale of the Self-efficacy in Technology and Science Instrument (SETS).

Scientific Inquiry and this Study: the *River City* World

This research is conducted within a technology-based curriculum, called *River City*. *River City* is a multi-user virtual environment (MUVE) designed to engage teams of two to four students in a collaborative scientific inquiry-based learning experience. In this world, students conduct their scientific investigations in a virtual historical town—populated by themselves, digitized historical artifacts, and computer agents—in order to detect and decipher a pattern of illness that is sweeping through the virtual community. Students manipulate a digital character, called an avatar, in order to explore the town; and they conduct virtual experiments to test their scientific hypotheses about the causes of the *River City* epidemic.

This research on self-efficacy was embedded in a larger, ongoing, NSF-funded project that has implemented *River City* nationwide with nearly 8,000 students since 2000. Previous research indicates that students are engaged by the virtual experimentation, that their scientific inquiry skills improve, and that their self-efficacy also increases (Ketelhut and Nelson, *in review*; Ketelhut *et al.*, *in press*).

In the larger study, in order to explore the effects of different pedagogical strategies on student motivation and learning, we designed several different *River City* treatments: one rooted in guided social constructivism, two rooted in different aspects of situated learning, and two versions that contain embedded guidance hints, termed the “high guidance” and “low guidance” treatments (Nelson *et al.*, 2005). Each implementation of the larger study assessed the impact of a sub-group of these different treatments, with students in the experiment being randomly assigned to them.

The guidance system embedded in *River City* was designed by Brian Nelson (2005) and offers constructivist hints to participants at various locations within the city. These hints promote reflection and offer scaffolds embedded in the context. Prompts for these hints, not the hints themselves, appear to students after they exhibit specific behaviors, such as entering a building. For example, a student entering the hospital and clicking on the admissions record for the hospital and then traveling to the tenement homes would be offered the opportunity to connect their gathered information. When activated, hint #2 would read: “*There are more mosquitoes here now. Are there more illnesses?*” The content of the hints are individ-

ually tailored to the participant, based on the cumulative history of his or her scientific exploration up to that point. The “high guidance” treatment offers three hints to a participating student each time it appears, whereas the “low guidance” treatment only offers one.

For the purposes of this exploratory study, I chose to investigate the scientific inquiry behaviors of students in a single treatment group—the “high guidance” group. By accessing the “high guidance” system, students had an additional source of information about the problems facing them as scientists in *River City*, and therefore, the record of their obtained behaviors in the City offers deeper insight into their decision process.

Scientific Inquiry Behaviors in River City

River City is a problem-based, student-centered project where students can gather evidence from the environment in diverse ways, based in part on the practices in which an epidemiologist might engage while investigating an outbreak of illness. For example, students are able to explore the town and gather tacit clues (e.g., about the topography of the town); they are also able to interview computerized residents, sample water and insects, visit the hospital, and look for clues in various other places they select, such as in embedded digital historical photographs and in the City library. In the City, students are guided to explore these various learning options, but not directed to choose specific sources of information or particular activities.

Multi-user virtual environments offer students a non-linear approach to learning. While various options may result in different kinds of learning, teams of students can succeed in solving the problem they are posed in *River City* using multiple alternate paths through a variety of sources that help develop their understanding. For example, one team of learners might choose to gather clues about the problem by interacting with computerized residents who describe their medical symptoms; another team might access the admissions record of the virtual hospital to see who was admitted, with what symptoms, and from what part of town. Because teamwork is strongly encouraged, the effect of these teams on individual student behavior was incorporated into the analysis.

For this research, I used the definition of scientific inquiry provided by the *National Science Education Standards* (National Research Council, 1996, p. 23) to identify the scientific behaviors that I recorded and analyzed for each participant. I list

them below, and I have mapped each onto the location in *River City* where the behavior could be observed:

- (1) “Making observations.” By moving around the world, students can make visual and auditory observations about the city and its inhabitants. A server-side database, through communication with the software, then records the student’s interactions and aggregates those into the student’s path for that exploration.
- (2) “Posing questions.” Students can also pose questions of the 32 computerized residents of *River City* and elicit short sets of information. Again, the database records what they ask and how the computerized resident answers their questions.
- (3) “Examining books and other sources of information to see what is already known.” Students can also access information directly from books in the *River City* library as well as from guidance hints, from embedded clues in digitized historical images, and from the hospital admissions records. Every time a student clicks his or her mouse on a source of information during a visit to *River City*, the time and identity of that source are recorded in the database.
- (4) “Using tools.” Students can also gather scientific data using two virtual microscope tools: a water sampling tool and a ‘bug-catching’ tool that are built into the software. Each tool is activated explicitly by a student mouse click that is then recorded in the database by name and with a timestamp.

Over the course of a 3-week implementation, students visited the *River City* environment on six separate occasions. During the first visit, students were primarily engaged in exploring *River City*, and the tasks that they performed focused on helping them become familiar with the software interface. Then, in each of the following three visits, students completed a new set of scientific mini-tasks designed to support the overarching goal of discovering the cause of the epidemic; each visit involved experiencing a different season in the virtual city (winter, spring, summer). These mini-tasks also helped introduce students to the tools available in *River City* for their investigation. During these three visits, participating students were focused on gathering information to help them formulate a scientific hypothesis in

response to the mystery they were posed. During the fifth and sixth visits, participating students were able to change one factor in one of two identical worlds, thus creating “control” and “experimental” worlds in which they tested their hypotheses about the source of the illness that was sweeping the City. In the analysis reported here, I focus on data from the students’ second, third and fourth visits to the City, as their behavior during these visits was devoted to scientific inquiry and was less circumscribed than during the first visit or the last two visits.

River City incorporates a server-side database that supports a wide variety of coding and analysis techniques. For example, an investigation of the interactions that students engage in with *River City*’s computerized residents can reveal to which residents students chose to talk, what they asked the residents, and whether the gender of the resident affected student choice of whom to interact with and what to say. Identifying, coding and analyzing student micro-behaviors at this level may cast important light on the interrelationship between students’ levels of self-efficacy and their information seeking behaviors over the course of the learning experience.

Specific Research Questions

Through its built-in database, *River City* offers data on how students experienced a novel, authentic scientific activity and provides an opportunity to investigate how student self-efficacy is related to scientific inquiry behavior, a hypothesized potential first link in the chain to improved learning in science. For this research, therefore, I explored the growth in scientific inquiry behaviors that individual students exhibited over their second, third, and fourth visits to the *River City* virtual world, by investigating the ways that they collected scientific evidence in the world. Then, I examined whether growth in inquiry behaviors differed by student self-efficacy in scientific inquiry, measured prior to entry into the virtual world.

In addition to looking at student scientific behaviors overall, I also examined growth in the diversity of sources from which individual students chose to gather data, and whether this growth too was related to self-efficacy in scientific inquiry. My specific research questions were therefore:

- (1) What growth in scientific inquiry behaviors overall do students exhibit in *River City*? Do students with lower self-efficacy gather less scientific evidence and demonstrate lower

growth rates of accessing scientific evidence than students with higher self-efficacy?

- (2) What growth in the diversity of sources for gathering data do students exhibit in *River City*? Do students with lower self-efficacy gather scientific evidence from fewer sources and demonstrate lower growth rates in diversity of sources than students with higher self-efficacy?

RESEARCH DESIGN

Site

I gathered data on a subset of students in 16 seventh-grade classes taught by four different teachers in one middle school in a public school system in New York State. These students represent the entire seventh grade in this district, and their teachers volunteered to implement the *River City* project in the context of their science classes. The student population in the district was approximately 80% white, with 3% eligible for free or reduced lunch (New York State, 2003).

Sample

I used data collected from the sample of 96 students who were randomly assigned individually to the “high guidance” treatment of the *River City* evaluation. As described previously, these students were chosen because they had access to an additional source of information in the guidance hints. This particular student sample was, like their school, somewhat homogenous: 6% were eligible for free or reduced lunch, 3% were categorized as special education (for emotional reasons), 11% spoke English as a second language. 52% of the students were male. The students showed considerable heterogeneity, however, in their entering scores. On a scale of one to five, student self-efficacy in scientific inquiry ranged from 1.8 to 4.8, with an average value of 3.5 and a standard deviation of 0.56. In addition, their pretest scores on scientific content ranged from 7 to 27 with an average of 16.3 (maximum possible score was 28) and a standard deviation of 5.

Procedures

- (1) Prior to implementation, students were randomly assigned within class to one of five treatments.

- (2) Teachers then created teams of 2–4 students within treatment and within class.
- (3) During the first two class periods of the *River City* implementation, students responded to two pre-surveys: one in which they self-reported on their affective characteristics (including the twelve-item *SETS* survey), and the second designed to assess their knowledge of disease and scientific inquiry, crafted specifically to evaluate the content to be taught in the *River City* intervention.
- (4) Students then spent approximately the next 10 days participating in the *River City* project, making a total of six visits with the remaining 4 days devoted to team design work and interpretation and whole class teacher-facilitated discussions. While in the City, their interactions and communications with the computer “residents” of the city were recorded in the server-side database. As indicated earlier, the data I analyze is drawn from students’ second, third, and fourth visits to *River City*, during which time their movements and choices of scientific behavior are primarily guided by their own interest. Based on classroom observations, the degree of team collaboration that took place during visits 2–4 differed widely by team; however, my analysis indicated that, for these three visits, team collaboration did not exert a significant effect on student behavior.

Measures

Outcomes

The two outcomes for my analyses were derived from measurements and coding of the scientific inquiry behaviors that students presented during each visit to *River City*; the behaviors measured repeatedly over the student’s second, third, and fourth visits. Since the goal of these behaviors is to access new information, repeat visits to the same source were only counted once, except for two of the variables, indicated below, where multiple visits to the same source could result in new information. In addition, I created thresholds of student involvement in the world; students who logged into a particular *River City* world for less than 15 min or engaged in less than 5 scientific inquiry behaviors during a single visit were counted as absent for that visit.

First, my outcome measures were derived from measurements on the following indicators:

- (1) The number of different places visited by each student during each visit per class period. This variable captures the *NSES*’ criterion of making observations (National Research Council, 1996, p. 23).
- (2) The number of different water sampling stations accessed by each student during each visit per class period. This variable measures the *NSES*’ criterion of using tools (National Research Council, 1996, p. 23).
- (3) The number of different “bug-catching” stations accessed by each student during each visit per class period. This variable measures the *NSES*’ criterion of using tools (National Research Council, 1996, p. 23).
- (4) The number of times that each student accessed the hospital admissions records during each visit per class period. This variable measures the *NSES*’ criterion of gathering evidence from book materials (National Research Council, 1996, p. 23).
- (5) The number of different digitized pictures with clues that each student clicked on during each visit per class period. This variable measures the *NSES*’ criterion of gathering evidence from book materials (National Research Council, 1996, p. 23).
- (6) The number of times each student accessed a library book during each visit per class period. This variable measures the *NSES*’ criterion of gathering evidence from book materials (National Research Council, 1996, p. 23).
- (7) The number of times each student interacted with different guidance messages in the individualized guidance system during each visit per class period. This variable measures the *NSES*’ criterion of accessing information from other sources (National Research Council, 1996, p. 23).
- (8) The number of different computerized agents of which each student asked ‘what’s new’ during each visit per class period. This variable measures the *NSES*’ criterion of posing questions (National Research Council, 1996, p. 23).

My two outcome variables were composites formed from the preceding indicators, as follows:

- (1) The total number of scientific inquiry behaviors engaged in by each student during each visit per class period, formed by adding all eight of the above indicators;
- (2) The diversity of choices made by each student in gathering evidence during each visit per class period. This variable was formed by counting the number of above categories of scientific inquiry behavior from which the student gathered data. The measure ranged from 0 to 8. During my growth modeling, I used the square root of diversity as my outcome, in order to ensure that the individual trajectories were linear in time.

Question Predictors

I was interested in seeing if time and initial self-efficacy predicted for the outcomes listed above. Time was represented by each student's second, third, and fourth visit to *River City*. Therefore, in my analysis, the "initial visit" referred to students' second visit to *River City*. I measured self-efficacy in doing inquiry science for each student prior to working in *River City* using the scientific inquiry subscale of the *SETS* instrument (Ketelhut, 2004). It contains 12 items; each rated on a scale from 1 (low) to 5 (high). Overall scores are computed by averaging the student's responses across the 12 items of the subscale, with high scores representing high self-efficacy. The measure has an estimated internal consistency reliability of 0.86 in a population of middle school students.

Control Predictors

In addition to these two question predictors, I controlled for students' prior knowledge, their gender and who their teacher was. The prior science knowledge of each student was measured on the pre-intervention assessment. This instrument consists of 28 multiple-choice questions on biology and scientific inquiry. Scores are computed by adding up the number of correct responses, and they range from 0 to 28 with higher scores representing more content knowledge. The measure has an estimated reliability of 0.86 (Ketelhut *et al.*, in press). I controlled for this variable in my regression analyses because Lawless and Kulikowish (1996) found that students' exploration patterns in hypermedia depended on their prior science knowledge.

Gender of each student was also controlled. Prior research on *River City* indicated that the relationship between use of the guidance system and content gain scores differed by gender (Nelson, 2005), and so, I controlled for this in each of my analyses. Lastly, I controlled for the fixed effect of the teacher of each student.

Data Analyses

Recall that my first research question is: What growth in scientific inquiry behaviors overall do students exhibit in *River City*? Do students with lower self-efficacy gather less scientific evidence and demonstrate lower growth rates of accessing scientific evidence than students with higher self-efficacy? To address this question, I conducted an individual growth modeling analysis of students' total scientific inquiry behaviors across the second, third, and fourth visits to *River City*, using SAS PROC MIXED, full maximum likelihood estimation (Singer and Willett, 2003). In my analyses, I hypothesized that individual student growth in total scientific inquiry behaviors was linear in visit (time). I included student self-efficacy score and my controls as predictors. Initially, I conducted a four-level analysis to account for the presence of time, student, the student-team, and the teacher, with the last being accounted for by the fixed effects of teacher. After accounting for time, student and teacher, I discovered that I could ignore the student-team, as its effect was not significant.

Since self-efficacy suggests that students with high self-efficacy will persevere longer and expend more effort (see previous review), I expected to find that students with higher self-efficacy engaged in more scientific inquiry behaviors than students with lower self-efficacy initially. Additionally, I expected to find that students with higher self-efficacy increased the number of scientific inquiry behaviors in which they engaged, across time, more rapidly than students with lower self-efficacy.

To address research question 2, I also used individual growth modeling to examine student changes over visits two, three, and four to *River City*. However, in this case, my outcome was replaced by the square root of my second outcome, an outcome that measured how varied were students' choices of where to gather scientific evidence. My analyses were similar to those conducted for research question 1 with the outcome replaced. I expected to find, similarly to research question 1, that students with higher self-efficacy chose to gather data from more varied

Table I. Parameter Estimates (Standard Errors) and Approximate *p*-Values for Fixed and Random Effects from a Series of Fitted Multi-level Models in Which the Two Outcomes are Predicted by Gender and Self-efficacy Over Time (*n* = 96)

	Outcome			
	Total number of behaviors			Diversity of data-gathering sources (square root)
	Model 1 Base model	Model 1a Self-efficacy	Model 1b Full model	Model 2 Base model
<i>Fixed effects</i>				
Intercept	11.50*** (1.35)	5.78 (3.75)	-1.46 (5.73)	1.60*** (0.06)
Visits	1.82** (0.58)	1.83** (0.58)	8.58* (3.60)	0.16*** (0.03)
Self-efficacy on entry		1.74 [‡] (1.07)	3.59* (1.58)	
Self-efficacy across visits			-1.64 [†] 0.98	
Male on entry			1.21 (1.76)	
Male across visits			-1.90 [†] (1.12)	
Teacher A	-1.06 (1.64)	-1.62 (1.65)	-1.65 (1.66)	-0.15* (0.07)
Teacher B	2.81 [†] (1.72)	2.51 (1.72)	2.85 (1.75)	-0.04 (0.08)
Teacher C	3.18* (1.65)	2.44 (1.69)	2.57 (1.71)	-0.01 (0.07)
<i>Random effects</i>				
Level-2 intercept	39.95*** (10.62)	38.44*** (10.46)	36.27*** (10.07)	0.06** (0.02)
Level-2 covariance	-14.88* (6.2)	-14.76* (6.15)	-12.52* (5.85)	-0.02 [†] (0.01)
Level-2 slope	12.55** (5.04)	12.87** (5.06)	10.67* (4.83)	0.03** (0.01)
Level-1 residual	22.82*** (4.08)	22.62*** (4.04)	22.78*** (4.06)	0.06*** (0.01)
<i>Goodness of fit statistics</i>				
-2LL	1568.2	1565.6	1559.9	142.4
Change in -2LL from previous model (Δdf)		2.6 [‡] (1)	5.7 (3)	
Change in -2LL from base (Δdf)		2.6 [‡] (1)	8.3 [†] (4)	
Pseudo <i>R</i> ² statistic	0.099	0.114	0.131	0.127

p* ≤ 0.05, *p* ≤ 0.01, ****p* ≤ 0.001, [†]*p* ≤ 0.10, [‡]*p* = 0.10.

sources than students with lower self-efficacy initially, and that their per visit rate of change would increase more rapidly than those with lower self-efficacy. This hypothesis stems from self-efficacy theory, which suggests that high self-efficacy students find complexity a challenge whereas low self-efficacy students see it as an obstacle.

FINDINGS

Table I presents fitted multi-level models to address both of my research questions. As noted in the preceding section, I fit these models using multi-level modeling that accounted for the nesting of time within student, student within team, and team within teacher. Teachers were represented by fixed effects. However, I found the effects of team to be negligible or non-existent in all models fitted, so I removed this level from my analyses.¹ Therefore, columns 2

through 4 outline the specifics of fitted multi-level models 1, 1a and 1b that address research question 1, while column 5 presents fitted multi-level model 2 that addresses research question 2. In the first eight rows, I list the estimated fixed effects on the outcome of each question and control predictor as well as any interactions present in that model. The next three rows list the teacher fixed effects. This is followed by a listing of the random effects in the next four rows. Lastly in the final four rows, I list goodness of fit statistics, including -2LL, its difference between models, and an estimated pseudo-*R*² statistic. In what follows, I use entries in the table to address my research questions.

Research Question 1: The Growth in Total Scientific Inquiry Behaviors

Table I presents three models developed to address research question 1. Model 1 and 1a, columns 2 and 3 of Table I, are interesting precursors to the final model. The first, Model 1, shows the effect of time on data gathering without regard to self-

¹This finding is possibly a result of the visits analyzed since team collaboration increases throughout the project culminating in the last week of the project, which was not a part of this analysis.



Fig. 1 Fitted total number of data-gathering behaviors as a function of visit, by self-efficacy ($n = 96$).

efficacy or gender, while the second shows the effect of time and self-efficacy without regard to gender. In the third column of Table I, Model 1b shows the final model that presents the full answer to research question 1.²

Controlling for teacher effects in Model 1, the unconditional growth model, I estimate that, initially on entry into *River City*, students conduct on average approximately 13 data-gathering behaviors, and they increase their total data-gathering behaviors by 1.82 behaviors per visit. Both effects are statistically significant ($p < 0.001$, $p < 0.01$, respectively). Model 1a addresses the second part of research question 1 concerning the effect of self-efficacy in scientific inquiry on the individual growth trajectories. In the fitted model, the main effect of self-efficacy indicates that students gather 1.74 more pieces of scientific data initially for every one point difference in self-efficacy, on average ($p = 0.10$). While this effect does not meet standard levels of statistical significance, it is interesting and suggestive in the context of this exploratory study. The rate of change in data gathering from visit to visit is the same for all students, regardless of their self-efficacy levels. Students increase their scientific data-gathering by 1.8 pieces of data per visit ($p < 0.01$). For example, a student entering in with high self-efficacy, at the 90th percentile (4.2, on the scale of 1–5), initially engaged in fourteen scientific data-gathering behaviors, while a student entering with low self-efficacy, at the 10th

percentile (2.9, on the scale of 1–5) initially engaged in two fewer behaviors. The rate of change in behavior of both students continued identically through the three visits.

As outlined in my research design section, I was interested in exploring whether students' content pretest score also impacted their scientific data-gathering behaviors. I found that this control predictor and self-efficacy were strongly correlated ($r = 0.52$), and therefore, when both predictors were included in the same model, neither was statistically significant as a consequence of collinearity. Therefore, I removed the content pretest score control predictor from the fitted models.

Model 1a demonstrates the impact of self-efficacy on scientific data gathering, but did gender affect this relationship as well? Model 1b of Table I includes student gender, self-efficacy and interactions of self-efficacy with visit. In this fitted model, self-efficacy impacts the initial level but only has a borderline effect on the rate of change of students' data-gathering behaviors ($p < 0.05$ and $p < 0.10$, respectively). The story is reversed for gender, with no effect initially, but a borderline effect on the rate of change of students' data-gathering behavior ($p > 0.10$ and $p < 0.10$, respectively).

As stated earlier, while the effect of self-efficacy and gender on rate of change is a borderline effect, this model will be analyzed cautiously due to the exploratory nature of this study. Since it is difficult to interpret the effects of this model directly, I will discuss them in the context of two extreme examples: that of high and low self-efficacy. In Figure 1, I display example fitted trajectories in scientific behavior

²Throughout this article, I have collapsed the effect of teacher into an average value.

over time, by gender and self-efficacy (set at 10% and 90%, for purposes of this example) for both boys and girls as predicted for by this model. Figure 1A illustrates the fitted trajectory for boys with high self-efficacy as well as low self-efficacy. Figure 1B parallels that figure, but for girls. Similar graphs could be drawn for any level of self-efficacy.

As can be seen in this figure, initially, there is a significant difference between the students with high and low self-efficacy ($p < 0.05$). Boys with high self-efficacy gather 16 pieces of data while boys with low self-efficacy collect only 11 pieces of data. A similar relationship is seen for the girls. The high self-efficacy girls accumulate nearly 15 pieces of evidence initially, while the low self-efficacy girls only 10. While there are small differences initially between boys and girls with similar self-efficacy, the effect of gender initially, as indicated earlier, is not statistically significant.

The growth trajectories for boys and girls show that the relative levels between students do not remain static. Boys with high self-efficacy show negligible rates of change, while those with low self-efficacy show positive rates of change ($p < 0.05$). The picture is somewhat different for girls. Girls with high self-efficacy, unlike the high self-efficacy boys, show a slightly positive rate of change ($p = 0.09$). Girls with low self-efficacy show the strongest positive rates of change ($p < 0.001$) of the four groups of students.

The result of these different growth rates can be seen in a comparison of visit two and visit four. During visit two, the only significant differences are between students with high self-efficacy and low; gender has only a small and insignificant impact on data-gathering initially. However, by visit four, this has reversed. Scientific data gathering during visit four is not affected by students' self-efficacy at all! The differences between differing levels of self-efficacy have converged to eliminate this as a predictor of data-gathering behavior. However, unlike in visit two, now gender has a borderline impact on scientific data-gathering. Girls, on average, gather 18 pieces of evidence while boys only collect 15 pieces ($p = 0.09$). Unlike the story that Model 1a told, once I control for gender, we see that students with low self-efficacy appear to engage in the world in such a manner as to diminish the initial effects of self-efficacy completely.

Research Question 2: The Growth in the Diversity of Sources for Gathering Data

The unconditional growth model, model 2 in the fifth column of Table I, addresses the first part of this

research question. Controlling for teacher fixed effects, students on average gather data from 2.5 out of the eight possible categories initially. Then, in subsequent visits, students enjoy a positive rate of change, increasing the number of sources accessed over time ($p < 0.001$). In no model, subsequent in the taxonomy of fitted models, was self-efficacy a statistically significant predictor. Therefore, I conclude that self-efficacy does not appear to affect the diversity of choices that students make as to where they gather their data, nor their growth in diversity across time.

CONCLUSION

In this study, I set out to explore students' trajectories of scientific investigation while participating in an inquiry-based science project, and to examine the role played by their self-efficacy in scientific inquiry in those patterns. I have several interesting findings.

Overall Growth

One of the concerns leveled at educational technology projects is that any findings of learning gains may be related to the technological novelty of the new intervention rather than something intrinsic to the pedagogy. While there is no doubt that novelty is appealing to children, if this were the only reason that students were engaged in *River City*, I would expect that students would exhibit diminishing returns over time. This is not what I found; instead, students increase their total scientific data-gathering behaviors by nearly two behaviors on average, on each visit to *River City*. While this appears to argue against the novelty-alone effect, recall that I have only examined student behavior in 3 visits. A new version of the *River City* intervention that is being designed currently will have students visiting the City 10 times for data-gathering only. I recommend that my study be repeated using this new version of the intervention to see if this effect is maintained over longer periods.

Self-efficacy

My results concerning the impact of student self-efficacy on learning is more complex. As discussed, self-efficacy researchers have found that students with high self-efficacy are more likely to expend additional effort, to see complexity as a challenge, and to diver-

sify their learning choices over students with lower self-efficacy (Pajares, 1995, 2000). Thus, I hypothesized that students with higher self-efficacy would tend to gather more scientific data from more sources than students with lower self-efficacy and that their scientific behaviors would increase faster over time. In this study, my conjecture turns out to be only partly true.

When looking at the initial effects of self-efficacy for randomly chosen students, I discovered that students with lower self-efficacy, as a group, do collect less scientific data than students with higher self-efficacy on entry into *River City*, in support of the literature. However, after three visits to the complex world of *River City*, students exhibit no differences in their scientific data-collection, based on self-efficacy. Thus, it seems as though the complexity of *River City* does not reinforce any differences that students bring with them, and perhaps even works to undo them. Furthermore, self-efficacy has no effect on the diversity of sources from which students collect their scientific data! This would seem to contradict the literature which suggests that low self-efficacy students should shy away from exploring the intricacies of the new world.

Why might this be so? One tenet of the theory of self-efficacy is that students develop a sense of self-efficacy based on their past experiences (Bandura, 1977). While I measured students' self-efficacy in scientific inquiry and then observed their behaviors of scientific inquiry, it is possible that students responded to the pre-experiment survey based on their experiences of scientific inquiry in the classroom where they either had, or did not have, success. However, I observed students' behavior in a MUVE. Is it possible that scientific inquiry self-efficacy expressed in the context of the classroom is quite different from scientific inquiry self-efficacy exhibited using a MUVE in a classroom? Perhaps students have a high self-efficacy for using a computer game-like technology that transfers to this environment overcoming their low self-efficacy in doing schoolwork? Or, is it possible that the motivation of using a MUVE in the classroom is so strong that it overcomes traditional effects of self-efficacy? If so, then this offers a possible way to help give all students success in science in school. While appealing, this phenomenon will need testing to see if it is replicable in the new longer *River City* curriculum and if so, what is the relationship among students' self-efficacy in scientific inquiry, schoolwork, and technology.

One other intriguing possibility is that the immersion as a scientist in *River City* helps students

modify their self-efficacy, which then results in patterns of growth that blur the differences initially resulting from differences in self-efficacy. Further testing to see if there is a relationship between exposure to data gathering and changes in students' self-efficacy is needed, as well as a way to test students' self-efficacy over time to see if changes in data-gathering are related to changes in self-efficacy.

Gender and Self-efficacy

The effect of gender, while interesting, presents only a borderline effect in this study. However, given the exploratory nature of this research, I report it as a relationship that needs further exploring. The effect of gender on initial levels of scientific data gathering was insignificant. Both boys and girls with high self-efficacy collected more data than both boys and girls with low self-efficacy, as discussed in the section above. However, differences in gender occurred when examining trends in scientific data gathering over time. Then, on average, girls ended up gathering more data than boys with little difference between high and low self-efficacy students, and only small differences between girls and boys. When considering gender, we see that girls with low self-efficacy have the greatest rates of change in scientific data gathering while boys with high self-efficacy have nearly flat growth trajectories.

This is a particularly intriguing pattern, as it is generally believed that girls are less motivated and less engaged by gaming technologies (Krotoski, 2004). With that in mind, we designed *River City* to have features that might appeal to girls, in particular. For example, there are more female residents of *River City* than male, and the mayor and university president are both women. Thus, I wonder whether this phenomenon that I have detected is particular to *River City*, due to these design features. Do girls with low self-efficacy participate in different scientific behaviors from other students? More investigation is needed to see if this effect is maintained with a larger sample of students and whether different subpopulations of students have different patterns of involvement.

LIMITATIONS

There are several areas that affected the outcomes of my study. First and foremost, because mine is an exploratory study, I cannot draw causal conclusions. Indeed, several of my findings are based

on significance levels of $p = 0.09$, higher than is traditionally used. Therefore, these conclusions should be treated cautiously until they are reproduced with other students.

Second, teachers are encouraged to implement *River City* as fits their teaching style and their students' needs. Therefore, no two teachers implement the intervention exactly the same way. Some teachers allow students to explore a particular world for more than one class period; others relinquish control to their students. This can, in an extreme example, result in students exploring the worlds out of order. To control for this threat to validity, as stated above, I controlled for the fixed effect of teacher, using dummy variables, and used students from 12 different classes of four different teachers. In addition, in my study, one teacher failed to have his students visit the second world. Therefore, all of his students were recorded as absent from that world, and possess only two waves of data in my analyses.

Third, it is possible that students' prior experience with computer games may have interacted with their ability to explore and gather data in *River City*. Since my analyses looked at an intervention that is placed within typical classrooms of students who possess variability in experiences and knowledge across a wide range of topics, I chose not to control for this potential effect. However, as discussed above, I recommend that future studies investigate whether these experiences, or self-efficacy in this arena, impact these outcomes.

Fourth, my analysis may have lacked statistical power. I possessed only three waves of data on the students in this study. In addition, 25% of the students only had two waves of data, due to missing their visit 2. As a result, these results should be treated cautiously until they are verified with a larger sample and more waves of data.

FINAL THOUGHTS

I started out in this study to understand more fully how differences in self-efficacy can affect students' participation in scientific inquiry. With advances in technology, I was able to follow students' moment-by-moment choices of behavior while gathering data. As is typical of exploratory research, I am left with more questions than answers—questions that in today's climate require investigating. While there is some indication that self-efficacy does effect data-gathering and that participating in a MUVE might undo that difference, this study has not looked

at whether data-gathering is related to learning outcomes or to changes in self-efficacy that can endure beyond its borders. Does low self-efficacy keep students from experiencing the wonders of science? If so, then we need to invest our time in figuring out how to raise the self-efficacy of students. There are hints in my analysis that embedding science inquiry curricula in novel platforms like a MUVE might act as this catalyst for change. Further research using these techniques will allow us to start to better understand the interaction between scientific inquiry and self-efficacy, and thus eventually science learning outcomes.

REFERENCES

- Alberts, B. (2000). Some thoughts of a scientist on inquiry. In J. Minstrell, & E. H. Zeevan (Eds.), *Inquiring into Inquiry Learning and Teaching in Science* (pp. 3–13). Washington DC: American Association for the Advancement of Science.
- American Association for the Advancement of Science. (1990). *Science for all Americans*, Oxford University Press, New York.
- American Association for the Advancement of Science (1993). *Project 2061: Benchmarks for science literacy*. New York: Oxford University Press.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review* 84(2): 191–215.
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*, Prentice Hall, Englewood Cliffs, NJ.
- Bong, M. (1996). *Perceived Similarity among Tasks and Generalizability of Academic Self-efficacy* (Reports—Research/Technical. Speeches, Conference Papers).
- Business Higher Education Forum (2005). *A commitment to America's future: Responding to the crisis in mathematics and science education*. Available: <http://www.bhef.com/MathEduReport-press.pdf> [2005, July].
- Coble, C., and Allen, M. (2005). Keeping America competitive: Five strategies to mathematics and science education. Denver: Education Commission of the States.
- Collins, J. L. (1984). *Self-efficacy and ability in achievement behavior*. Stanford University, Unpublished thesis.
- Grigg, W., Lauko, M., and Brockway, D. (2006). *The Nation's Report Card: Science 2005* (NCES 2006-466). US Department of Education, National Center for Education Statistics. Washington, DC: US Government Printing Office.
- Jinks, J., & Morgan, V. (1996). Students' sense of academic efficacy and achievement in science: A useful new direction for research regarding scientific literacy. *Electronic Journal of Science Education* 1(2).
- Ketelhut, D. J. (2004) *Assessing science self-efficacy in a virtual environment: A measurement pilot*. Harvard University, Cambridge, MA, Unpublished qualifying paper.
- Ketelhut, D. J., and Nelson, B. (in review) Designing for real-world inquiry in virtual environments.
- Ketelhut, D. J., Dede, C., Clarke, J., and Nelson, B. Studying situated learning in a multi-user virtual environment. In E. Baker, J. Dickieson, W. Wulfecck, & H. O'Neil (Eds.), *Assessment of Problem Solving Using Simulations*. Lawrence Erlbaum Associates, Mahwah, NJ (in press).
- Krotoski, A. (2004). *Chicks and joysticks: An exploration of women and gaming*. Entertainment and Leisure Software Publishers

- Association. Available: http://www.elspa.com/assets/files/c/chicksandjoysticksanexplorationofwomenandgaming_176.pdf [2006, July 1].
- Lawless, K. A., and Kulikowich, J. M. (1996). Understanding hypertext navigation through cluster analysis. *Journal of Educational Computing Research* 14(4): 385–399.
- Lemke, J. (1990). *Talking science: Language, learning, and values*, Ablex Publishing, Norwood, NJ.
- Lent, R. W., Brown, S. D., and Larkin, K. C. (1984). Relation of self-efficacy expectations to academic achievement and persistence. *Journal of Counseling Psychology* 31: 356–362.
- Leonard, W. H., Speziale, B. J., and Penick, J. E. (2001). Performance assessment of a standards-based high school biology curriculum. *American Biology Teacher* 63(5): 310–316.
- Lopez, F. G., and Lent, R. W. (1992). Sources of mathematics self-efficacy in high school students. *The Career Development Quarterly* 41: 3–12.
- National Research Council. (1996). *National science education standards: Observe, interact, change, learn*, National Academy Press, Washington, DC.
- Nelson, B. (2005). *Investigating the impact of individualized, reflective guidance on student learning in an educational multi-user virtual environment*. Harvard University, Cambridge, MA, Unpublished dissertation.
- Nelson, B., Ketelhut, D. J., Clarke, J., Bowman, C., and Dede, C. (2005). Design-based research strategies for developing a scientific inquiry curriculum in a multi-user virtual environment. *Educational Technology* 45(1): 21–27.
- New York State (2003). *New York state school report card comprehensive information report*. Available: <http://www.emsc.nysed.gov/reprcdfall2003/cir/662401060008.pdf> [2005, July].
- Pajares, F. (1995). *Self-efficacy in academic settings*. Paper presented at the American Educational Research Association, San Francisco, CA.
- Pajares, F. (1996). *Assessing self-efficacy beliefs and academic outcomes: The case for specificity and correspondence*, [Internet]. Available: <http://www.emory.edu/EDUCATION/mfp/aera2.html> [2004, February 21].
- Pajares, F. (1997). Current directions in self-efficacy research. In M. Maehr & P. R. Pintrich (Eds.), *Advances in Motivation and Achievement* (Vol. 10, pp. 1–49) Greenwich, CT: JAI Press.
- Pajares, F. (2000). *Schooling in America: Myths, mixed messages, and good intentions* (Lecture): Emory University, Cannon Chapel.
- Pajares, F., and Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology* 86(2): 193–203.
- Singer, J., and Willett, J. (2003). *Applied longitudinal data analysis*, Oxford University Press, New York.
- Smith, P. L., and Fouad, N. A. (1999). Subject-matter specificity of self-efficacy, outcome expectancies, interests, and goals: Implications for the social-cognitive model. *Journal of Counseling Psychology* 46(4): 461–471.
- Tai, R., Liu, C., Maltese, A., and Fan, X. (2006). CAREER CHOICE: Enhanced: Planning early for careers in science. *Science* 312(5777): 1143–1144.
- Zimmerman, B. J., and Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal* 31: 845–862.